UNIVERSITÀ DEGLI STUDI DI PISA
Facoltà di Scienze Matematiche Fisiche e Naturali
Corso di Laurea Specialistica in Scienze Fisiche
Anno Accademico 2002-2003

Tesi di Laurea Specialistica

Correlations and Hierarchies in the Spanish Stock Market

Candidato:
Mirko Stefano MEGA

Relatore esterno:
Dott. Esteban Moro Egido

Relatore interno:
Dott. Riccardo Mannella

To my parents.

Contents

Introduction

Since the 1900s, the economists work of analysing and modelling the financial market has been supported by pioneering groups of mathematicians, starting with the work of Bachelier [1]. These collaborations have been very fruitful, so much that nowadays a robust theoretical framework characterises these disciplines. It is only since the 1990s, after the developments by physicists of many useful tools and methodologies for the analysis of complex systems, that a group of physicists started some parallel studies of economic systems as dynamical systems. This delay is related to earlier difficulties for the physicists to speak about "predictability" in systems which do not have a microscopic description; and to the fact that the unpredictability of human behaviour had been taken as a basic key factor for the impossibility to apply physical methods to describe and explain the market evolution.

During the last 30 years, physicists have achieved important results in the field of phase transitions, statistical mechanics, disordered systems, and non linear dynamics. These results have made the physical predictions less strong and quite different from the meaning originally associated to Newtonian dynamics. On the other hand, the area of research covered by physical investigation has increased. Indeed, similarities between financial markets and hierarchical systems of interacting particles may be an interesting way to enhance the synergy between social and physical sciences.

Nowadays the contribution by physical research is well present in many different disciplines, from medicine to biology, from social networks to finance, and so on. This is the proof that the physical spirit of an "experimental approach", analysing the empirical data, extrapolating the main characteristics of the system, and building a kind of theoretical model that explains them, is the correct spirit.

Financial markets, with their power-law distributions, correlations, unpredictable time series, scaling, and random processes, exhibit several of the properties that characterise complex systems. They are open systems in which many little units interact nonlinearly in the presence of feedback but with a subtle difference: if scale invariance meant the absence of characteristic scale to a physicist, from a business point of view it would mean the existence of catastrophic risk, which can bankrupt a company.

Most likely, the financial market is the most monitored system of the human history. With the help of electronic storing of data, a huge amount of financial data is currently available, useful to test new models. The available information in the stock market is of three levels: the first level contains only the information about the prices of stocks at each time, the second level contains in addition to the first, more information about the buy/sell orders, and at last, the third level contains all of the above information plus the code of who makes orders. While Level I information is usually free, high-frequency Level III information is expensive and difficult to obtain. Our study refers to the Level III information and constitutes one of the first studies of high frequency data in terms of agents.

The aim of this thesis is to analyse the main general properties of a particular financial market, using some of the latest techniques developed by physicist in different fields, as for example: social network methods for the study of interactions among agents, random matrix methods and correlation time series methods for interaction between stocks, and distribution analysis for time series. The bottom line of these attempts bears on the basic principle, allowing for a real progress in any scientific model, which can be resumed in the following strategy:

> *"First you guess. Don't laugh, this is the most important step. Then you compute the consequences. Compare the consequences to experience. If it disagrees with experience, the guess is wrong. In this simple statement is the key to science."*
>
> Richard P. Feynmann.

The analysis of the market is done from the point of view of describing the system as a statistical system, pulling out some well known economic properties with the help of physical tools and building a market model which might reproduce the main statistical properties.

While doing so, we will be more interested, for instance, in understanding if the market obeys some general hypothesis (like the efficient market hypothesis, see page 18) which are normally assumed to be satisfied, or same breakdown is observed. And if a breakdown is observed , we are interested in understanding why: it is in other words the underlying dynamics which is interesting for us. The bottom line is an attempt to obtain some feeling about the agent-agent interaction. Thus, our aim is quite far from predicting future market movements by identifying and exploiting its inefficiencies (which nowadays many physicists converted into market players pursue as the Holy Grail).

The structure of this thesis is the following:

In the first chapter the financial market is described in its general form, with a brief introduction about its history and general characteristics, in particular the Spanish financial market, subject of our applications. Then, a comparison with more general complex systems is done, focusing the attention on the common properties of these two kinds of systems.

In the second chapter the analysis of the real data begins, starting from the analysis of the statistical properties of shares prices. Following the same line of earlier analysis done for the New York Stock Exchange (NYSE), the purpose here is to compare some of the main properties of the Spanish market with the ones of NYSE, to prove empirically that they belong to the same class of markets and suggest that we might have some universality at work. This is a fundamental step to generalise the new results obtained here to other stock markets. At the end, the study of the correlation among shares is carried out, by using concepts like ultrametrics, to choose the shares that are most correlated and to group them together in one "megashare", to be used in the study of correlations among agents.

In the third chapter the analysis shifts to the study of the characteristics and of the correlations among agents in the Spanish market in the attempt to build some description of their interaction. Here, it is shown that the distribution of wealth of the agents in the Spanish stock market obeys a Pareto law, which reflects the observed distribution of wealth in most western societies. New variables, characterising the information contained in an agent network, are defined for the purpose of normalising the different enthusiasm of agent depending on time of the day. These variables are used to investigate, with techniques of social networks analysis, possible dependences or hierarchies. These results are then compared with the ones of a random network showing that the homogenous market participants hypothesis, assumed by the *efficient market hypothesis*, is fallacious.

In the fourth chapter some recent market models, such as spin models and agent-based models, are described along with their merits and shortcomings. Then an agent-based model, a modification of the Cont-Bouchaud model, is proposed. This model introduces a herding behaviour among agents, modelled according to the agents hierarchical structure found in the third chapter. Thus the model is used to build price time series with statistical properties similar to the ones of real price time series, proving that the supply of external information does not affect those statistical properties.

In the last chapter, the conclusions of this analysis are drawn. The Spanish stock market has the same statistical properties of other financial markets already studied. Its price returns distributions exhibit inverse power law tails, decaying with the same exponents of the S&P500 index of NYSE. There is also a well defined hierarchical structure among stocks composing the Ibex-35® index. The analysis of agents shows that the agents are heterogenous and a hierarchical structure exists among them. These two aspects are at variance with the assumptions of the *efficient market hypothesis*, and with the assumption which many market models do, that the participants of the market are homogenous and share the same information, and that is all the information is available in the market. This disagreement could be the cause of the observed power law in the return distributions of real data, as shown by the proposed model.